

# PATHWAYS TO SOCIAL EVOLUTION: RECIPROCITY, RELATEDNESS, AND SYNERGY

Jeremy Van Cleve<sup>1</sup>

Erol Akçay<sup>2</sup>

**Key words:** cooperation, non-additive fitness, behavioral responses, genetic assortment

<sup>1</sup> Santa Fe Institute  
1399 Hyde Park Road  
Santa Fe, NM 87501 USA  
e-mail: [vancleve@santafe.edu](mailto:vancleve@santafe.edu)

<sup>2</sup> Department of Ecology and Evolutionary Biology  
Princeton University  
Guyot Hall  
Princeton, NJ, 08544 USA  
e-mail: [eakcay@princeton.edu](mailto:eakcay@princeton.edu)

*Date modified: 13<sup>th</sup> November, 2012*

## **Abstract**

Many organisms live in populations structured by class and location, exhibit plastic responses to their environment and their social partners, and are subject to non-additive ecological and fitness effects. Social evolution theory has long recognized that all of these factors can lead to different selection pressures but has only recently attempted to synthesize how these factors interact. Using models for both discrete and continuous phenotypes, we show that analyzing these factors into a consistent framework reveals that they generically interact with one another in ways not previously emphasized or recognized. Specifically, behavioral responses, genetic assortment, and synergy all interact in non-trivial ways that cannot be summarized by simple summary indices. In contrast, some recent attempts at synthesis yield flawed conclusions because they neglect these interactions, and thus lead to an incomplete picture of how natural selection shapes social phenotypes. We suggest that these interactions should be a boon for empirical research as they are likely crucial for describing the complexity of biological systems.

## 1. Introduction

Explaining the evolution of social behaviors has been a goal of evolutionary theory going back to Darwin’s time. The modern effort to explain the origin of social behaviors both cooperative and noncooperative started with the seminal work of Hamilton [1, 2], which showed that genetic relatedness has a profound influence on evolutionary dynamics. Later work added other significant factors, starting with conditional and responsive behaviors [3, 4] and extending to non-additive fitness interactions [5]. Beginning in the 1980’s, researchers analyzed many models that apply and extend these mechanisms [e.g., extending reciprocity to include indirect responses in ref 6]. At the same time, empirical researchers tested the role of various mechanisms developed in the theoretical literature [e.g., 7–9].

Whereas early discussions of social evolution usually pitted one mechanism against another, recent efforts are increasingly focused on synthesizing and integrating the various pathways of social evolution [e.g., 10–15]. This is perhaps the surest sign that social evolution theory has entered a mature phase. These efforts at synthesis are not always without controversy, but they have served to illuminate the connections between various models devised over the decades. In this paper, we aim to contribute to this synthetic effort by focusing specifically on the relation between three important concepts: assortment of genotypes (relatedness), behavioral responsiveness (e.g., reciprocity), and non-additive (e.g., synergistic) effects of social behaviors on fitness. These topics have of course been written about extensively in the past; yet, how these concepts relate to one another is still not well understood.

Two of the most fundamental mechanisms for the evolution of cooperation are responsive (and conditional) behaviors such as direct and indirect reciprocity, and genetic assortment that creates relatedness between individuals [11]. Both factors are clearly in operation in many, if not most social interactions. Cooperative breeding, for example, almost always involves interactions with relatives of varying degree. At the same time, most cooperatively breeding species also exhibit conditional behavior; for example, non-helpers might get evicted from the breeding group [16]. Likewise, reciprocal food-sharing among vampire bats can occur amongst relatives as well as non-relatives [17].

Non-additive interactions between individual phenotypes also play a prominent role in social

evolution. Sometimes, the non-additivity is a direct function of organismal metabolism, as in the case of the plant pathogen *Agrobacterium tumefaciens*, which induces infected hosts to produce opines, which can be used as a resource only by those cells that also carry the virulence plasmid [18]. In other cases, the ecological context might create non-additive payoffs. For example, in the penduline tits (*Remiz pendulous*), males contribute to the building of an elaborate nest, while females lay eggs inside the nest [19]. After eggs are laid, one or both of the parents desert the nest, which is then tended either by a single parent or by neither parent. Data on breeding success suggests that the reason for this behavior is in part that a single parent has better than even chance of raising the clutch [20]. Hence, the second parent adds less to breeding success than the first parent, meaning a negative non-additivity. It has been argued that the opposite is true for many bird species with biparental care [21]. It is worth noting that in both of these examples, as perhaps with most instances of non-additive interactions between phenotypes, the traits in question are plastic. For example, the *Agrobacterium* pathways inducing opine production are not expressed constitutively and instead depend on quorum-sensing [18]. Likewise, the male and female decisions to desert in penduline tits appear context-dependent; the same individual can care or desert according to factors such as brood size [19, 22].

Yet, despite enormous theoretical work on conditional behaviors, genetic assortment, and non-additive payoffs, the interplay between these three factors has attracted relatively little attention until recently [11, 14, 23–25]. Some of the recent work synthesizing these different factors makes strong claims about the relationships between them. For example, Fletcher & Doebeli [14] asserts that the fundamental condition favoring cooperation is assortment between cooperative genotypes and cooperatively behaving partners, regardless of whether the assortment is caused by conditional behaviors or population structure. Other authors, notably Queller [5, 15] and Fletcher & Zwick [23], claimed that reciprocity is inherently a multiplicative or non-additive interaction between phenotypes. However, as we show below, both of these claims rely on models of limited generality and do not hold in general. Our aim is therefore to sort out the exact relationships between behavioral responses, genetic assortment, and non-additivity.

Our paper begins with two models: first, we introduce a discrete-action model to clarify the relationship between genetic assortment, behavioral responses, and synergistic or non-additive interactions; second, we generalize the results of the discrete-action model by presenting a population

genetic model for the evolution of a continuous trait in a structured population, where the structure can be due to any combination of age, stage, sex, or spatial location. Using these models, we show that the claims that reciprocity and genetic assortment are simply equivalent mechanisms generating phenotypic assortment [14] and that their effects contribute additively to the response to selection [15] are generally incorrect. Instead, these factors have independent effects that can interact in non-linear ways. We also show that synergistic interactions, while important, are not necessary for capturing the effect behavioral responses and reciprocity. Rather, synergistic interactions can affect the stability of social behaviors or even change the payoff structure of the game entirely.

## 2. Discrete action model

### *2.1. Reciprocity and genetic assortment*

Our model begins with a 2x2 game that is played repeatedly between two individuals a large number of times. We take the game to be symmetric for ease of exposition, but this assumption can be relaxed with no consequence for our main argument. Label the two actions  $C$  and  $D$ . Most models of such repeated games focus on particular conditional strategies (e.g., tit-for-tat [4] and its numerous relatives). We would like a behavioral model that admits a wider range of propensities to respond to one's opponent. Hence, we assume that each individual  $i$  has two traits:  $a_i$ , which is the action ( $C$  or  $D$ ) that the individual  $i$  intrinsically "prefers", and  $\rho_i$ , the probability that individual  $i$  overrides its intrinsic preference and reciprocates its opponent's action in the last round. Hence, in common parlance,  $a_i$  corresponds to whether individuals are cooperators or defectors, and  $\rho_i$  denotes the degree to which individuals are reciprocators. Both traits are genetically encoded; we take  $a_i$  to be binary and  $\rho_i$  to be a continuous trait between 0 and 1. As an aside, a straightforward extension of this model could treat  $a_i$  as a continuous trait by interpreting it as the intrinsic probability of playing  $C$  vs.  $D$ .

Consider an interaction between an intrinsic cooperator and defector (i.e.,  $a_1 = C$  and  $a_2 = D$ ). Suppose at stage  $t$ , player 1 (the  $C$ -type) played  $C$  and player 2 (the  $D$ -type) also played  $C$  (because he was reciprocating cooperation in the previous round). We label this situation as  $(C, C)$ , where the elements correspond to the first and second players' actions, respectively. Let us label the response coefficients of the two individuals as  $\rho_c$  and  $\rho_d$  in according with intrinsic preferences for

$C$  and  $D$ , respectively. In that case, the probability that at stage  $t+1$  they will keep playing  $(C, C)$  is the probability that the  $D$ -type individual keeps reciprocating, i.e.,  $\rho_d$ . Similarly, the probability that the  $t+1$  stage will result in  $(C, D)$  is  $1-\rho_d$ , as this is the probability that the  $D$ -type will revert to his intrinsic preference. Finally, the probabilities of playing  $(D, C)$  and  $(D, D)$  at  $t+1$  given the actions at  $t$  is zero, since regardless of whether she plays her intrinsic preference or reciprocates, the  $C$ -type player will always play  $C$ . In a similar way, we can work out the probabilities of playing any action pair at  $t+1$  conditional on the action pair at  $t$ , which is summarized with the transition matrix  $\mathbf{M} = [m_{ij}]$ ,

$$\mathbf{M} = \begin{pmatrix} \rho_d & (1-\rho_d) & 0 & 0 \\ (1-\rho_c)\rho_d & (1-\rho_c)(1-\rho_d) & \rho_c\rho_d & \rho_c(1-\rho_d) \\ 0 & 1 & 0 & 0 \\ 0 & 1-\rho_c & 0 & \rho_c \end{pmatrix}, \quad (1)$$

where  $m_{ij}$  gives the probability of being in state  $j$  at  $t+1$  when the game was in state  $i$  at  $t$ , and states are ordered as  $\{(C, C), (C, D), (D, C), (D, D)\}$ . We assume that the game is played for a very large number of periods with no discounting, so that the proportion of time each pair of actions is played is given by the stationary distribution of the Markov chain given by the transition matrix  $\mathbf{M}$ . Taking the dominant left-eigenvector of  $\mathbf{M}$  and normalizing it, we find the stationary state distribution of the game between a  $C$ - and  $D$ -type individual,  $\pi$  to be

$$\pi = ((1-\rho_c)^2\rho_d, (1-\rho_c)(1-\rho_d), (1-\rho_c)\rho_c(1-\rho_d)\rho_d, \rho_c(1-\rho_d)^2) (1-\rho_c\rho_d)^{-2}. \quad (2)$$

The expected payoff of a  $C$ -type player from playing  $D$ -type player is the fraction of time spent at each outcome (given by the elements of  $\pi$ ) multiplied by the payoff to her from that outcome. Suppose that at each period the game has the familiar structure of the prisoner's dilemma (in the other-only benefit form [26]): cooperation yields a benefit of  $b$  to the partner, at a net cost of  $c$  to the cooperator. Defection carries no cost and yields no benefit. Thus, the expected payoff to a

$C$ -type player when playing a  $D$ -type,  $w_{CD}$ , is given by

$$w_{CD} = \frac{(1 - \rho_c)(b\rho_d - c)}{1 - \rho_c\rho_d}. \quad (3)$$

With a similar calculation, we find the expected payoff to  $D$ -type from playing a  $C$ -type to be

$$w_{DC} = \frac{(1 - \rho_c)(b - c\rho_d)}{1 - \rho_c\rho_d}. \quad (4)$$

The expected payoff to a  $C$ -type from playing another  $C$ -type is trivial; since both players start with  $(C, C)$  and keep playing these actions,  $w_{CC} = b - c$ . Similar reasoning yields  $w_{DD} = 0$ .

In order to determine which type of individual is preferred by natural selection, we calculate the condition when the frequency of intrinsic cooperators with responsiveness  $\rho_c$  increases in a population with intrinsic defectors of responsiveness  $\rho_d$ . Suppose that some mechanism (e.g., kin-recognition, population viscosity, or greenbeards) causes cooperators to assort themselves so that an intrinsic cooperator interacts with another intrinsic cooperator with probability  $r_c$ ; likewise, intrinsic defectors assort among themselves with probability  $r_d$ . When not assorting specifically with their own types, individuals are paired with a partner randomly chosen from the population (this is analogous to the classic assumption concerning populations with related individuals; see equation 1 in [27]). A full description of the dynamic of frequency change in the population would require a specification of the mechanism that generates assortment, and how it depends on the frequency of genotypes. However, we can leave this unspecified if we only require a qualitative picture of frequency change over the course of a single generation.

Given our assumptions above and that the frequency of the intrinsic cooperative genotype is  $q$ , the expected fitnesses of an intrinsic cooperator and an intrinsic defector are, respectively,

$$\begin{aligned} W_C &= (r_c + (1 - r_c)q)w_{CC} + (1 - r_c)(1 - q)w_{CD} \quad \text{and} \\ W_D &= (1 - r_d)qw_{DC} + (r_d + (1 - r_d)(1 - q))w_{DD}. \end{aligned} \quad (5)$$

The condition for increase of intrinsic cooperators is  $W_C - W_D > 0$ , which leads to

$$\frac{b}{c} > \frac{1 - (1 - q)\rho_c - q\rho_d + (1 - q)r_c\rho_c(1 - \rho_d) + qr_d(1 - \rho_c)\rho_d}{q\rho_c + (1 - q)\rho_d + (1 - q)r_c(1 - \rho_d) + qr_d(1 - \rho_c) - \rho_c\rho_d}. \quad (6)$$

What this equation entails can be seen better by looking at three special cases. First, suppose individuals do not respond to each other,  $\rho_c = \rho_d = 0$ , and that the assortment probabilities are equal,  $r_c = r_d = r$ . In this case, condition (6) becomes  $\frac{b}{c} > \frac{1}{r}$ , which of course is Hamilton's rule [1]. Second, suppose both the intrinsic cooperators and defectors have the same responsiveness,  $\rho_c = \rho_d = \rho$ , and assortment is zero,  $r_c = r_d = 0$ . Then, condition (6) evaluates to  $\frac{b}{c} > \frac{1}{\rho}$ , which is equivalent to other rules for the evolution of responsiveness [28] and reciprocity [11, 29]. Finally, giving intrinsic cooperators and defectors the same assortment probability and responsiveness,  $r_c = r_d = r$  and  $\rho_c = \rho_d = \rho$ , yields

$$\frac{b}{c} > \frac{1 + r\rho}{r + \rho}. \quad (7)$$

In condition (7), the assortment coefficient  $r$  and responsiveness  $\rho$  have symmetric effects on the increase condition but are also separate because both the product and the sum of  $\rho$  and  $r$  appear. Hence, assortment of individuals and conditional behaviors cannot be collapsed to a single index, as Fletcher & Doebeli [14] have suggested. Not coincidentally, expression (7) is identical in structure to the condition for an evolutionary increase in cooperation we have previously derived for continuous public goods games [25] and to other previously derived results [11, 24]. This further reinforces the point that this basic symmetry and irreducibility is not due to a quirk in this particular model.

## 2.2. *Reciprocity and synergism*

We can use the same model to investigate the interplay between reciprocity and synergism, the latter defined as an extra benefit (or cost, for negative synergism) when two cooperators interact. Since Queller's seminal paper in 1985 [5], synergism, or non-additive interactions in general, appears in several efforts to synthesize social evolution theory [e.g. 15, 23, 30]. Fletcher & Zwick [23], for example, consider synergistic interactions as the consequence of reciprocal altruism, and thus propose that Queller's 1985 rule should serve as the basis of integration for these historically distinct explanations for cooperation. We show in the next section that reciprocity need not imply non-additivity in models of continuous traits. In this section, we show that reciprocity and synergism in phenotypes (i.e., some non-additive payoff to cooperative behavior) have distinct effects on the dynamics of social traits.

As is customary in models of synergism, we modify the prisoner's dilemma game matrix above

by adding a constant  $d$  to the payoffs of both individuals when they both cooperate. When  $d > 0$ , two cooperators produce more benefit together than twice the benefit of a single cooperator; when  $d < 0$ , they produce less. These cases relate to cooperative acts by the partners being strategic complements versus substitutes, respectively, as defined in economics [31].

Using the same methodology above, we calculate the expected payoff to an intrinsic cooperator when paired with an intrinsic defector and an intrinsic defector against an intrinsic cooperator as

$$\begin{aligned} w_{CD} &= \frac{(1 - \rho_c) ((b\rho_d - c)(1 - \rho_c\rho_d) + d\rho_d(1 - \rho_c))}{(1 - \rho_c\rho_d)^2} \quad \text{and} \\ w_{DC} &= \frac{(1 - \rho_c) ((b - \rho_dc)(1 - \rho_c\rho_d) + d\rho_d(1 - \rho_c))}{(1 - \rho_c\rho_d)^2}, \end{aligned} \quad (8)$$

respectively. These expressions indicate that as with reciprocity and population structure, reciprocity and payoff synergism also interact in more complex ways than recognized. To better illustrate the interaction, suppose that both types of players have the same probability of reciprocation,  $\rho_c = \rho_d = \rho$ , and the same probability of assortment,  $r_c = r_d = r$ . Then, as shown in Appendix A, the component that the payoff synergy  $d$  adds to the condition for cooperation to increase,  $W_C - W_D > 0$ , is

$$d \left( p + r(1 - p) + \frac{\rho(1 - 2p)(1 - r)}{(1 + \rho)^2} \right), \quad (9)$$

where the first component,  $p + r(1 - p)$  matches previous work on synergy without reciprocity [32, 33]. The first thing to note from (9) is that, compared the non-synergistic case in condition (7), synergy adds frequency dependence to the increase condition. This is a well-known result [33–35] that complicates the analysis of synergy. Essentially, this frequency dependence arises from explicitly accounting for the frequency of different pairings, which is necessary when payoffs are not additive. Another way of thinking about this frequency dependence is that while relatedness or genetic assortment generally accounts for pairwise genetic correlations, accounting for synergy requires calculating triplet correlations, which are generally much more tedious to calculate [35, for example]. In models with continuous phenotypes, assuming that the two competing phenotypes are very similar (i.e., weak phenotypic effect; 27, 36, 37), the effect of non-additive payoffs are neglected to first order in the difference between the phenotypes (defined as  $\delta$  in section 3). However, as we will see below in section 3, synergy can still have an effect in terms of which phenotypes are stable

in the long term.

Second, the evolutionary game, defined by the expected payoffs from the repeated game,  $w_{CC}$ ,  $w_{CD}$ ,  $w_{DC}$ , and  $w_{DD}$ , will take on different game structures depending on the values of  $d$  and  $\rho$ . Figure 1 illustrates the different possibilities. With  $\rho < c/b$ , a high enough synergism can transform the game from a prisoner’s dilemma into a coordination game, which will have two evolutionarily stable strategies (ESSs): all cooperators or all defectors. If  $\rho > 0$ , then still higher  $d$  can turn the game into a mutualism game, where intrinsic cooperators have an unconditional advantage. When  $\rho > c/b$ , the possibility of generating a coordination game disappears, and increasing  $d$  turns the prisoner’s dilemma into an anti-coordination game (i.e., a Hawk-Dove or Snowdrift game) and eventually into a mutualism game as  $d$  increases further. It is well known that reciprocity in the form of repeated game strategies such as tit-for-tat can transform the payoff matrix [38–40, for example], but to our knowledge, the interaction between reciprocity and synergy demonstrated in Figure 1 has not been recognized. This interaction is particularly important in the light of recent results that show that genetic non-additivity by itself does not change the conditions for the evolution of cooperation [35], unless it changes the structure of the evolutionary game into a coordination or anti-coordination game, which have different equilibrium structures and produce different patterns of frequency dependence.

### 3. Population genetic model

While the discrete-action model above highlights the interplay between genetic assortment, behavioral responsiveness, and phenotypic non-additivity, it is a restrictive model. A more general population genetic framework will allow us to generalize the above results to interactions among an arbitrary number of individuals who may be from different demographic classes (e.g., males and females, queens and workers, juveniles and adults, philopatric and dispersed). A population genetic model will also show us how synergistic interactions that result from nonlinear ecological benefits and costs can be crucial in the long-term evolutionary stability of phenotypes. The study of such synergies as they relate continuously to ecological payoffs is relatively new [e.g., 25, 28, 41] and presents an opportunity to make predictions about the effects of specific ecological mechanisms on the evolution of specific behaviors. At the same time, such generality involves a significant increase in the complexity of the model, and hence a different set of simplifying assumptions come into play.

In particular, we are interested in determining when a mutant allele that codes for some phenotype with social effects will, on average, increase in frequency in a population and reach fixation. The proper quantity to calculate in this case is the fixation probability. For structured populations, the fixation probability may be impossible to obtain analytically, so we follow a standard approach that approximates the that probability using a first-order a Taylor-series expansion in the difference between the mutant and resident phenotypes,  $\delta$  [27, 42]. Population genetics theory shows that the first-order term in the expansion is sufficient to calculate the standard conditions for the evolutionary stability of phenotypes [27, 42]. Moreover, this first-order term is proportional to a quantity that is much easier to calculate, namely the derivative of the expected change in the frequency of the mutant allele with respect to  $\delta$  or  $d\Delta q/d\delta$ . It is this derivative that we calculate below.

First, we will write a general expression for  $\Delta q$  in a class structured model using standard theory in population genetics and evolutionary demography [42–45]. We do this using an “individually” centered approach that was popularized in population genetics by the Price equation [46, 47]. Suppose that there are  $K$  classes in the population, where an individual  $k$ ’s class could represent the local population it lives in, its age or developmental stage, its sex, or some other phenotypic feature. Each class contains  $N_k$  individuals and the total population size is  $N = \sum_k N_k$ . In order to track allele frequencies in different classes, we let  $\mathbf{q} = (\dots, q_{jk}, \dots)$  be a column vector of length  $N$  where  $q_{jk}$  is the frequency of the allele in individual  $k$  in class  $j$ . Average allele frequencies in each class in the next time period are calculated by multiplying a transition matrix  $\mathbf{W} = [w_{ijk}]$  times  $\mathbf{q}$ , where  $w_{ijk}$  is the probability that an allele in a random individual in class  $i$  descended from individual  $k$  in class  $j$ . The transition matrix is a function of the fitnesses of each class and the transmission processes that affect how alleles move between classes such as segregation, mutation, and migration. We assume that the fitness of individual  $k$  in class  $j$  is function of its phenotype,  $p_{jk}$ . In the simplest cases, phenotype depends on genotype in a linear way where  $p_{jk} = p + q_{jk}\delta$  and  $\delta$  is a measure of how much the mutant differs from the resident phenotype  $p$ . For non-additive genetic interactions (such as dominance or epistasis), phenotype depends on genotype in a nonlinear way [34].

To calculate  $\Delta q$ , we need to average the class frequencies in such a way that the average frequency remains constant in the absence of natural selection. The appropriate weights for this

average are the reproductive values of each class [43, 48, 49],  $\alpha = (\dots, \alpha_k, \dots)$ . Thus, the change in the reproductive value weighted allele frequency is

$$\Delta q = \alpha \cdot \mathbf{W} \cdot \mathbf{q} - \alpha \cdot \mathbf{N} \cdot \mathbf{q} \quad (10)$$

where  $\mathbf{N}$  is a  $K \times N$  matrix used to create class averages of the allele frequencies  $\mathbf{q}$  and whose  $k$ -th row has the value  $1/N_k$  in the columns  $1 + \sum_{l=1}^{k-1} N_l$  to  $\sum_{l=1}^k N_l$  and zero elsewhere for all  $k$ . Using equation (10), the derivative of  $\Delta q$  with respect to  $\delta$  is given by

$$\frac{d\Delta q}{d\delta} = \alpha \cdot \frac{d\mathbf{W}}{d\delta} \cdot \mathbf{q} = \sum_{i,j,l=1}^K \sum_{k=1}^{N_j} \sum_{m=1}^{N_l} \alpha_i \frac{\partial w_{ijk}}{\partial p_{lm}} q_{jk} \quad (11)$$

The components of equation (11) are reproductive value,  $\alpha_i$ , the effect of one individual's phenotype on another individual's probability of transmitting an allele from one generation to the next,  $\frac{\partial w_{ijk}}{\partial p_{lm}}$ , and the probability that those two individuals both have the mutant allele,  $q_{jk}q_{lm}$ . When this last component,  $q_{jk}q_{lm}$ , is averaged over all individuals in a class, we obtain a genetic identity probability that can be related to identity by descent probabilities from classic population genetics [27, 42, 50] that emerge as more common quantities such as  $F_{ST}$  [51] in simple models of population subdivision [50] such as the island model [52]. Moreover, these genetic identity probabilities correspond to the genetic covariance terms found in regression models in social evolution [53], which when normalized by genetic variance correspond to the regression definition of relatedness [54, 55].

Although equation (11) allows us to see how evolutionary change should generally rely on genetic covariances to first order, understanding the interplay between ecological and genetic mechanisms requires unpacking the derivatives of the transition matrix,  $\frac{\partial w_{ijk}}{\partial p_{lm}}$ . Evaluating these derivatives can be very tricky if the demographic or class structure affects the social interaction since we must in effect track the fertility, demographic, and genetic parameters of each class separately. Thus, we will assume a relatively simple social interaction that is unaffected by the class structure. Assume that each individual in each class participates in a social interaction with  $n - 1$  other individuals from the same class (e.g., males interact with males, females with females, adults with adults, etc). Each individual choose actions in this social interaction (e.g., an amount of effort in some helping behavior); we denote the behavioral equilibrium [25, 28] or time averaged values of the action of

individual  $k$  in class  $j$  by  $a_{jk}^*$ . One crucial quantity of interest is the marginal change in the action of one individual in response to another; this is the response coefficient  $\rho$  from section 2.1, which can be written as  $\rho_{j,kl} = \frac{\partial a_{jl}^*}{\partial a_{jk}^*}$  for the change in the action of individual  $l$  in response to a change in the action of  $k$ , both of class  $j$  [25, 28]. Since we assume that the social interaction is unaffected by class, we can neglect subscripts that specify class in variables involved in the social interaction; thus, the response coefficient can be written as  $\rho_{j,kl} = \rho_{kl}$ .

The actions individuals choose in a social interaction are determined by some proximate behavioral mechanism whose biological components (e.g., neurological structures) are genetically determined (whereas actions themselves are not inherited). For simplicity, we assume that a single “motivational” phenotype  $p$  characterizes the proximate mechanism, and it is the evolution of this phenotype that we track with equation (10). One way of measuring how the motivational phenotype affects actions is  $\frac{\partial a_l^*}{\partial p_k}$ , which is the effect of individual  $k$ ’s phenotype on the action of  $l$ . The response coefficient can be expressed as a function of these derivatives:  $\rho_{kl} = \frac{\partial a_l^*}{\partial a_k^*} = \frac{\partial a_l^*}{\partial p_k} / \frac{\partial a_k^*}{\partial p_k}$ .

The outcome of the social interaction determines the fertility of each individual in the interaction ( $f_{jk}$  for individual  $k$  in class  $j$ ); for example, an individual’s fertility might be determined by the net calories it obtains from participating in group hunting and foraging activities. Since the social interaction does not depend on class, fertilities are independent of class ( $f_{jk} = f(p_{jk})$ ). We can define the “costs” and “benefits” of the social interaction in terms of fertility as  $-c = \frac{\partial f}{\partial a_{\bullet}^*}$ , where “ $\bullet$ ” represents the focal individual, and  $b = \frac{\partial f}{\partial a_{\circ}^*}$  for some other individual “ $\circ$ ”. Given these assumptions, we show in Appendix B that  $d\Delta q/d\delta$  is

$$\frac{d\Delta q}{d\delta} \propto b\rho(n-1) - c + \mathcal{R} (n-1) [b(1 + \rho(n-2)) - \rho c] = \mathcal{S}, \quad (12)$$

where  $\rho$  is the response coefficient in a resident population monomorphic for the phenotype  $p$  and  $b$  and  $c$  are evaluated at the resident phenotypic value  $p$ . Equation (12) is the condition for increase of the phenotype  $p$  to first order in the phenotypic deviation, and  $\mathcal{S} > 0$  is the  $n$ -player analog of inequality (7). This result was previously derived by Akçay & Van Cleve [25] and is analogous to expressions found in McGlothlin *et al.* [24] and Lehmann & Keller [11]. It is important to note that equation (12) is essentially a rearrangement of Hamilton’s rule [1] where the effect of kin competition is moved from the benefit and cost terms into the *scaled* relatedness-coefficient  $\mathcal{R}$

[42, 50, 56, 57]. We derive an expression for  $\mathcal{R}$  in Appendix B.

The first thing to note about equation (12) is that it incorporates our measure of reciprocity, the responsiveness coefficient  $\rho$ , without requiring a non-additive genetic model, contrary to previous suggestions that non-additive interactions must be the basis of integrating reciprocity with genetic assortment [5, 15, 23]. This implies that, to first order in the strength of selection, the effect of reciprocity is not frequency dependent even though reciprocity may still generally produce frequency-dependent selection. One reason to emphasize this distinction is an implicit assumption in the literature that frequency dependent selection implies the need to consider second-order and higher terms in Taylor expansions of the change in allele frequency [e.g. 58]. Secondly, equation (12) shows that relatedness  $r$  and responsiveness  $\rho$  cannot be both replaced by a single index of “assortment” [14], which reinforces the point made in section 2.1 that relatedness and responsiveness are distinct causal factors that interact synergistically.

If we are interested in what social behaviors or phenotypes evolve in the long-term, we can set  $\Delta q = 0$  in equation (11) and solve for candidate behaviors or phenotypic values that are evolutionarily stable. Setting  $\Delta q = 0$  is equivalent to setting the expression in (12) equal to zero (i.e.,  $\mathcal{S} = 0$ ). In order to ensure that these candidate phenotypes are actually stable, we need to check whether populations can evolve or converge towards these candidate phenotypes starting from neighboring phenotypes [36, 59, 60]. This kind of evolutionary stability, called convergence stability, is often more important in the long-term evolution of continuous phenotypes than the traditional notion of strict evolutionary stability [61].

A candidate phenotypic value is convergence stable when the derivative of the  $d\Delta q/d\delta$  with respect to the resident phenotypic value  $p$  is negative [36]. This condition amounts to

$$\frac{d}{dp} \left[ \frac{d\Delta q}{d\delta} \right] \propto \frac{d\rho}{dp} \frac{\partial \mathcal{S}}{\partial \rho} + \frac{db}{dp} \frac{\partial \mathcal{S}}{\partial b} + \frac{dc}{dp} \frac{\partial \mathcal{S}}{\partial c} < 0, \quad (13)$$

where the derivatives in (13) are *total* derivatives and measure how  $\rho$ ,  $b$ , and  $c$  change due to both direct and indirect changes in the phenotype  $p$ . Even in this generic form, condition (13) tell us two important things about the what influences the long-term evolutionary stability of social behaviors. First, evolutionary stability depends on how a change in phenotype affects the level of responsiveness,  $\frac{d\rho}{dp}$ , and how responsiveness changes selection for that phenotype,  $\frac{\partial \mathcal{S}}{\partial \rho}$ . Thus, it is

not only the level of reciprocity that matters, but the sensitivity of reciprocity levels to changes in the underlying phenotype.

Second, condition (13) says that stability depends on how changes in the phenotype result in changes in the benefits and costs,  $\frac{db}{dp}$  and  $\frac{dc}{dp}$ , respectively. Recall that  $b$  and  $c$  are defined in terms of derivatives of fertility with respect to actions, so  $\frac{db}{dp}$  and  $\frac{dc}{dp}$  generate second derivatives of fertility respect to the actions of individuals. These second derivatives of fertility roughly describe the economics of how the behaviors of individuals combine to produce offspring. For example, the derivative of a focal individual  $k$ 's fertility with respect to its own action and the action of another individual  $l$ ,  $\frac{\partial^2 f_k}{\partial a_k^* \partial a_l^*}$ , measures the economic complementarity or substitutability of the actions of two different individuals. When this derivative is positive, actions are complementary and increasing effort by one individual increases the fertility benefit of effort by another individual. A negative derivative indicates substitutable actions and the reverse effect on a fertility benefit. Complementary actions can be thought of as positive ecological synergies, which can have very significant effects on the evolution of cooperative behaviors [25, 28, 41]. Effectively, non-zero values of these second derivatives indicate that payoffs as a function of phenotype are non-additive. However, our increase condition for the mutant allele did not only accounts for first order effects and thus does not include non-additive effects. Thus, in a model with weak effect mutants, the potential ESS phenotypes are unaffected by non-additivity, whereas the stability of those phenotypes is affected.

## 4. Applications

### 4.1. Assortment and reciprocity

How does the general condition in (6) compare to the condition that Fletcher & Doebeli [14] derive (their equation 2.2) using a single index for assortment? Fletcher & Doebeli argue that we need to compare the average interaction environment for cooperators and defectors, which will depend not only on the interaction probabilities,  $r_c$  and  $r_d$ , but also how much each genotype ends up playing C and D against intrinsic cooperators and defectors, respectively. For intrinsic cooperators, the frequency of stage games where the opponent plays C, which we denote by  $e_c$  to match Fletcher and Doebeli's notation, is given by:  $e_c = r_c + (1 - r_c)(q + (1 - q)(\pi_1 + \pi_3))$ , where the last term is simply the sum of the fraction of games between a C- and D-type where the outcome is (C, C) and (D, C) (calculated at the stationary distribution). Similarly,  $e_d = (1 - r_d)q(\pi_1 + \pi_2)$ . These

coefficients now capture Fletcher and Doebeli’s notion of the average interaction environment. But if we followed Fletcher and Doebeli’s procedure, we would arrive at the following condition <sup>1</sup>:

$$\frac{b}{c} > \frac{1}{e_c - e_d} = \frac{1 - \rho_c \rho_d}{r_c(1 - \rho_d)(1 - q) + r_d q(1 - \rho_c) + \rho_d(1 - \rho_c) + q(\rho_c - \rho_d)}, \quad (14)$$

Comparing this expression with (6), we can see that it is not the correct condition for the increase of cooperators. Why does Fletcher and Doebeli’s method fail? It fails simply because reciprocity and other types of behavioral responses affect a genotype’s fitness not only through changing the social environment a genotype experiences, but also through changing the phenotype expressed by the same genotype in response to this social environment. Fletcher and Doebeli’s method of partitioning selection into effects due to own phenotype and others’ phenotypes remains valid, but assortment in their analysis only affects the others’ phenotypes part of the partition. In contrast, behavioral responses affect the focal individual’s own phenotype as well. Therefore, the assertion that behavioral responses and non-random interactions are only two special cases of the same fundamental quantity, assortment, does not hold. Rather, both factors have distinct (albeit symmetric) effects on the selection pressure on social behaviors.

Incidentally, Fletcher and Doebeli were not the first ones to make this erroneous claim. In an influential paper, Queller [5] states that his derivation of the Hamilton’s rule accounting for multiplicative effects is valid regardless of whether the the covariance between the focal individual’s genotype is caused by assortment of genotypes or conditional behaviors. However, as we show above, this statement is only true if only one these factors are in place. In the presence of both genetic assortment and behavioral responses, the covariance between focal genotype and partner phenotype is not enough to account for the selection pressure, since the covariance between the focal genotype and the focal phenotype is altered as well. Below, we show that a more recent formulation of social selection by Queller [15] suffers from a different but related problem.

#### 4.2. *Separating kin and kith*

We now use the tools derived above to reconsider one of the recent and more prominent attempts to disentangle the different causal pathways for shaping the evolution of social traits. Queller [15]

---

<sup>1</sup>Fletcher and Doebeli use the “total benefit” parametrization of the public goods game, as opposed to our “other-only” definitions, so we also translate our  $b$  and  $c$  to match theirs.

provides an “expanded Hamilton’s rule” given by the following inequality [15, equation 1]

$$-c + \sum_i b_i r_i + \sum_i d_i s_i + \sum_i m_i f_i > 0, \quad (15)$$

where  $c$ , as usual, is the cost of a cooperative act to the focal individual,  $b_i$  is the benefit accruing to the  $i$ -th relative of the focal individual,  $r_i$  the relatedness coefficient between the focal individual and the  $i$ -th relative. The third and fourth terms represent an expansion of the classical Hamilton’s rule to include multiplicative effects and behavioral feedbacks, which Queller calls “kind” and “kith” selection, respectively. Of these terms, kind selection ( $\sum_i d_i s_i$ ) involves a coefficient quantifying the deviation from additivity  $d_i$  for the  $i$ -th interaction partner, which is a “synergistic” fitness effect, and a synergy regression coefficient  $s_i$ . Finally, the last term ( $\sum_i f_i m_i$ ) is composed of a coefficient measuring how the  $i$ -th social partner’s phenotype responds to the focal individual’s phenotype  $f_i$ , times the fitness effect of this response  $m_i$ . Queller suggests that this expansion of the phenotypic version of Hamilton’s rule represents more types of fitness effects in a causal manner than the classical form, and therefore might lead to better disentangling the various pathways through which selection acts on social behaviors. While we agree with the basic premise, we show below that equation (15) lacks a term reflecting the interaction between relatedness and behavioral feedbacks and is therefore correct only under restrictive assumptions.

To see this, we simplify equation (15) by dropping the “kind” selection component (i.e. setting  $d = 0$ ) and obtain

$$-c + \sum_i b_i r_i + \sum_i m_i f_i > 0. \quad (16)$$

The equivalent increase condition using our population genetic model is  $\mathcal{S} > 0$  from equation (12) for pairwise social interactions ( $n = 2$ ), which yields

$$-c(1 + \rho r) + b(r + \rho) > 0 \quad (17)$$

for the case where the behavioral phenotype affects of the fertility of all classes equally ( $i = 1$  in equation 16). The discrete action model yields an equivalent result in expression (7). In our model,  $\rho$  measures behavioral responsiveness, hence it should be analogous to the kith selection feedback coefficient,  $f$ , of Queller (specifically,  $\rho$  should be a function of  $f$ ). This allows us to

compare equations (16) and (17), which reveals that equation (17) has an additional term,  $-r\rho c$ , not included in equation (16); this means that a generalized Hamilton’s rule that includes both kin and kith selection cannot be constructed by simply adding the kin and kith components together.

The problem in equation (15) stems from Queller’s assumption, without any explicit justification, that the behavioral feedbacks and indirect fitness effects can be decomposed additively. This is apparent in the description of his equation (1), where he asserts that the behavioral feedback terms simply arise from decomposing direct fitness effects (i.e., the effect of the focal individual’s behavior on its own fitness) into non-social and social components. This assumption is correct only under very restrictive circumstances, namely when the only individuals that respond to the focal individual are those with zero relatedness to the focal individual, whereas related individuals do not respond. In other words, there are two classes of social partners that exert selection on the same behavioral trait: related conspecific partners that do not exert kith effects but do have kin effects and unrelated or heterospecific partners that do exert kith effects. In that case, the increase condition becomes (see Appendix C for derivation)

$$-c + b(\rho + r) > 0 ,$$

which is equivalent to the additive combination of kin and kith components in equation (16). However, in most interactions among social partners with responsive phenotypes, both related and unrelated individuals will exhibit behavioral feedbacks. In that case, as (12) shows, both direct and indirect fitness effects will be functions of behavioral responses ( $\rho$ ), and Queller’s additive separation of kin and kith effects becomes incorrect.

## 5. Discussion

One of the major goals of evolutionary and behavioral ecology is to elucidate the casual biological pathways that drive the evolution of social traits. However, natural selection rarely operates through a single pathway, and therefore the various causal components of social evolution need to be integrated and their commonalities and interactions explored. Our main goal in this paper is to contribute to this synthesis effort. In so doing, we highlight that genetic assortment, behavioral responses, and non-additive interactions between phenotypes all have both distinct and interacting effects. Therefore, combining these components requires an explicit and consistent modeling

approach. While some of our results (like those relating relatedness and reciprocity) are found in previous literature [11, 24, 25, 62], their general implications are not widely appreciated, or in some cases, there are misleading claims about these implications. The results that we emphasize in this paper are as follows.

(i) Phenotypic assortment is not a univariate quantity: it matters how much assortment comes from both behavioral responses and genetic assortment. Further, behavioral responses and genetic assortment cannot simply added to each other, even when interactions are additive. Rather, they interact with each other both additively and multiplicatively. Prior results analogous to this can be found in Lehmann & Keller [11], McGlothlin *et al.* [24] and Akçay & Van Cleve [25]. Yet, despite this interaction being known for some time, its crucial role seems not widely appreciated. Akçay & Van Cleve [25] is the first paper to our knowledge to highlight the synergistic and symmetric effect of genetic assortment and behavioral responses on the response to selection of social phenotypes. In that work, we demonstrated how synergism between genetic assortment and behavioral responses can facilitate the transfer of fitness effects from the within-group level to the between-group level, a crucial step in evolutionary transitions in individuality [63]. These previous results assumed continuous phenotypes; here we show that a similar synergism is found in interactions with discrete phenotypes.

(ii) Behavioral responses and reciprocity do not require non-additive genetic or phenotypic interactions, despite assertions to the contrary suggestions that measuring reciprocity requires non-additive terms due to frequency dependent selection [5, 15, 23]. This is a specific example of a more general confusion regarding the connection between frequency dependent selection and weak selection approximations of evolutionary dynamics in structured populations that generate expressions such as (12) that are frequency independent. In general, behavioral responses and reciprocity will generate both additive and non-additive interactions where the additive interactions can be thought of as the average effect of reciprocity and non-additive interactions contribute higher order statistical moments. Including only additive interactions and capturing only the average effect of reciprocity still generates expression (12), which captures the intuitive and crucial notion that reciprocity correlates phenotypes and thus can lead to increasing levels of cooperation.

(iii) Although measuring behavioral responses and reciprocity does not require including non-additive interactions, non-additive interactions can still be important, even in models with weak

selection and continuously varying phenotypes. Specifically, non-additive payoffs or fitness will affect the evolutionary (convergence) stability of a phenotype and how the stable phenotype changes with underlying parameters. The former effect can be seen in equation (13), which shows how non-additivity in the benefits and costs of a social phenotype (behavior) appear in the convergence stability condition. Non-additivity in benefits and costs relate directly to ecological mechanisms that convert the effort of individuals in the social interaction into payoffs in terms of fertility or survival. For example, individuals who put effort into a group hunt (which might be modeled by a stag-hunt game [39]) may receive more return in terms of the amount of prey item per unit of caloric investment when more individuals join the hunt. This kind of ecological synergy, or “complementarity” in economics [31], can have large effects on the level of investment in cooperative interactions both within groups [28] and in structured populations [25, 41]. It is important to note that synergies in discrete models can also model complementarity; for example, the positive values of the parameter  $d$  in expression (9) can also represent complementary actions (and negative values represent the opposite, called “substitutable” actions), though non-additive interactions generate frequency dependence in such models.

### ***5.1. Approximate and exact models of evolutionary change***

Fundamentally, there are two main mathematical approaches to modeling the effect of selection on continuous traits. The first approach, and the one that we employ, uses a first-order Taylor series to approximate the effect of selection when it is weak and the mutant and resident phenotypes are similar (i.e., small  $\delta$ ). This approach is justified by the utility of the first-order term in calculating the evolutionary stability of specific phenotypes [27, 42]. However, the first-order approximation will not describe the evolutionary dynamics very well when selection is strong and second order and higher effects matter. In contrast, the second approach [38, 48, 53, 64–66] fits a *linear* regression to the full change in allele frequency or mean phenotype. The result is that the linear regression represents the *exact partial* change in allele frequency or mean phenotype due to natural selection [53, 67, 68]. The generality of such an approach comes, however, at the cost of a “clear separation between genetic and ecological parameters” [p. 24 in ref 35]. For example, the partial change approach makes it impossible to tell when ecological mechanisms generate frequency dependent selection that might result in the maintenance of phenotypic variation. Also, mechanistic non-

additive interactions, such as synergistic payoffs [34] and epistasis [69], become buried within linear regression coefficient, which makes it difficult to tell when such interactions are important. We feel that the regression approach is a useful tool for qualitatively describing evolutionary change, while the Taylor series approach is better suited for studying the mechanisms underlying the evolution of social behaviors.

## 5.2. *Empirical issues*

The empirical study of the interplay between behavioral responses and genetic assortment and the role of synergistic payoffs is still relatively new. Most of the work on the linkage between behavioral responses and genetic assortment comes from the now rapidly growing application of indirect genetic effect (IGE) models to laboratory, agricultural, and field-based populations [e.g. 70–79]. IGE models [24, 80, 81] extend classical quantitative genetics [82] to include a coefficient,  $\psi$ , that measures the effect of a social partner’s phenotype on the phenotype of a focal individual. Importantly, the IGE coefficient  $\psi$  directly maps to our behavioral responsiveness coefficient  $\rho$  [25] and is thus also a measure of reciprocity. We know from equation (12) (and analogously equation 18 in McGlothlin *et al.* 24) that an interaction between behavioral responses and relatedness is important for selection on social traits, even when no such interaction is assumed to affect individual phenotypes (see equation C.5 in the Appendix or equation 5 in ref [24]). However, most of the empirical work to date spends little time on the interaction of behavioral responses and relatedness with the exception of Bleakley & Brodie [75] and Frère *et al.* [77], and only Frère *et al.* [77] emphasize the crucial role of this interaction. The empirical work assumes individual phenotypes to have an explicit interaction term, and it is not yet clear how empirical estimates of this interaction can be tied to predicted interaction term that arises in (12) that measure the effect of selection on the trait. Thus, more work needs to be done understanding the interaction between behavioral responses and relatedness both theoretically and empirically.

Empirical studies explicitly focused on non-additive payoffs seem scarce, except in the context of green-beard genes [83, e.g.]. We do not believe that this state of affairs reflects the importance of non-additive interactions between phenotypes in nature. Instead, non-additive and non-linear interactions are likely to be abundant and their effects strong. For example in parental care, many arguments implicitly or explicitly rely on the contributions each parent combining non-

additively [28, 84]. We suggest that measuring non-additive interactions at both the payoff (fertility or survival) and fitness levels will prove important for explaining the evolution of social behaviors, regardless of whether the phenotypes in question vary more or less continuously or are discrete characters with large effects.

To conclude, a careful analysis of behavioral responses, genetic assortment, and non-additive interactions shows a complex interaction between these three pathways. We believe that focusing on this complexity will ultimately deliver a more complete and nuanced understanding of the evolutionary forces shaping social behaviors. The necessary mathematical frameworks for such understanding exists, though as we show in this paper, it needs to be applied with some care. Once crafted, these theories can be tested against the growing abundance of sophisticated datasets that measure both networks of social interactions as well as important demographic and fitness-related variables [e.g. 85, 86].

## Acknowledgements

J.V. was supported by a Santa Fe Institute Omidyar Fellowship. E.A. was supported by National Science Foundation Grant EF-1137894, to S.A. Levin.

## Appendix A: Two-player game with synergy

We follow the discrete-action model defined in section 2.1. Individuals are either intrinsic cooperators or intrinsic defectors who can either play their intrinsic strategy or reciprocate their partner's last action. Let  $\rho_c$  and  $\rho_d$  be the probabilities that an intrinsic cooperator and defector, respectively, reciprocate or copy their partner's last action. We allow for non-randomly interaction among intrinsic cooperators and intrinsic defectors that might results from population structure or other assortment mechanisms and define  $r_c$  and  $r_d$  to be the probabilities that intrinsic cooperators and defectors, respectively, interact among themselves non-randomly. Two cooperators who play each other each earn payoff  $b - c + d$ , where  $b$  is the benefit,  $c$  is the cost of cooperation, and  $d$  is a synergy term that account for non-additive payoffs. A cooperator and defector playing together results in a payoff of  $-c$  for the cooperator and  $b$  for the defector. Two defectors playing each other each earn zero payoff. Given this setup and following the analysis of section 2.1, the payoffs to an intrinsic cooperator paired with an intrinsic defector and vice-versa are

$$w_{CD} = \frac{(1 - \rho_c) ((b\rho_d - c)(1 - \rho_c\rho_d) + d\rho_d(1 - \rho_c))}{(1 - \rho_c\rho_d)^2} \quad \text{and}$$

$$w_{DC} = \frac{(1 - \rho_c) ((b - \rho_dc)(1 - \rho_c\rho_d) + d\rho_d(1 - \rho_c))}{(1 - \rho_c\rho_d)^2},$$

respectively, which are also given in equation (8) in the main text.

We can plug the above values for  $w_{CD}$  and  $w_{DC}$ , along with  $w_{CC} = b - c + d$  and  $w_{DD} = 0$ , into the expression for fitness in equation (5) to derive the condition for cooperation to increase,  $W_C - W_D > 0$ , which simplifies to

$$\begin{aligned} & b \left( \frac{p\rho_c + (1 - p)\rho_d + r_c(1 - p)(1 - \rho_d) + r_dp(1 - \rho_c) - \rho_c\rho_d}{1 - \rho_c\rho_d} \right) \\ & - c \left( \frac{1 - (1 - p)\rho_c - p\rho_d + (1 - p)r_c\rho_c(1 - \rho_d) + r_dp\rho_d(1 - \rho_c)}{1 - \rho_c\rho_d} \right) \\ & + d \left( \frac{(p + r_c(1 - p))(1 - \rho_d)(1 - \rho_c^2\rho_d) + (1 - p + pr_d)(1 - \rho_c)^2\rho_d}{(1 - \rho_c\rho_d)^2} \right) > 0. \end{aligned} \quad (\text{A.1})$$

This expression simplifies to condition (6) when  $d = 0$ . Another simplification is to assume that the reciprocation and assortment probabilities are the same for cooperators and defectors, namely

$\rho_c = \rho_d = \rho$  and  $r_c = r_d = r$ . In this case, expression (A.1) simplifies to

$$b \frac{r + \rho}{1 + \rho} - c \frac{1 + r\rho}{1 + \rho} + d \left( (p + r(1 - p)) + \frac{\rho(1 - 2p)(1 - r)}{(1 + \rho)^2} \right) > 0, \quad (\text{A.2})$$

which in turn simplifies to condition (7) when  $d = 0$ .

## Appendix B: Allele frequency change in a class structured model

In section 3, we build a general class-structured model of allele frequency change and present a first-order approximation, expression (12), that is valid given the phenotype of one individual can affect the fertility of another individual so long as both individuals are in the same class and the effect of phenotype on fertility is independent of class. Our goal in this section is to show how to arrive at expression (12) beginning with

$$\frac{d\Delta q}{d\delta} = \alpha \cdot \frac{d\mathbf{W}}{d\delta} \cdot \mathbf{q} = \sum_{i,j,l=1}^K \sum_{k=1}^{N_j} \sum_{m=1}^{N_l} \alpha_i \frac{\partial w_{ijk}}{\partial p_{lm}} q_{jk} q_{lm}. \quad (\text{B.1})$$

Recall that the Taylor series is expanded around  $\delta = 0$ , so the partial derivative  $\frac{\partial w_{ijk}}{\partial p_{lm}}$  in (B.1) are evaluated at neutrality where natural selection on fertility differences is absent. Additionally,  $\delta = 0$  implies that all individuals in a class are equivalent, which means that the partial derivatives above fall into three cases: the first case is the effect of a focal individual's phenotype on its own probability of sending an allele to class  $j$  in the next generation from its own class  $i$ ,  $\frac{\partial w_{ij}}{\partial p_{j\bullet}}$ ; the second is the effect on that probability of another individual's phenotype from the same class as the focal,  $\frac{\partial w_{ij}}{\partial p_{j\circ}}$ ; and the third is the effect on that probability of the phenotype of an individual in a different class than the focal individual,  $\frac{\partial w_{ij}}{\partial p_l}$  where  $l \neq j$ . Substituting these derivatives into (B.1), we obtain

$$\alpha \cdot \frac{d\mathbf{W}}{d\delta} \cdot \mathbf{q} = \sum_{i,j=1}^K \alpha_i N_j \left( \frac{\partial w_{ij}}{\partial p_{j\bullet}} \overline{q_{j\bullet}^2} + (N_j - 1) \frac{\partial w_{ij}}{\partial p_{j\circ}} \overline{q_{j\bullet} q_{j\circ}} + \sum_{l \neq j} N_l \frac{\partial w_{ij}}{\partial p_l} \overline{q_j q_l} \right). \quad (\text{B.2})$$

Further, we assume that the phenotype only affects fertility (though it could affect survival or demographic factors such as carrying capacity or dispersal rates), which means the derivatives in the left-hand side of the equations in (B.3) can be split into two components, the effect of the phenotype on fertility and the effect of fertility on the probability an allele in class  $i$  descends from class  $j$ . The latter derivatives we represent with  $\tau$  where  $\tau_{ij,j\bullet} = \frac{\partial w_{ij}}{\partial f_{ij\bullet}}$ ,  $\tau_{ij,j\circ} = \frac{\partial w_{ij}}{\partial f_{ij\circ}}$ , and  $\tau_{ij,l} = \frac{\partial w_{ij}}{\partial f_{il}}$  for  $l \neq j$ . These derivatives essentially contain transmission probabilities due to dispersal, mutation, allelic segregation, age-dependent mortality, etc, and we assume they are independent of phenotype.

Given these two classes of derivatives,

$$\frac{\partial w_{ijk}}{\partial p_{lm}} = \begin{cases} \frac{\partial w_{ij}}{\partial p_{j\bullet}} = \tau_{ij,j\bullet} \frac{\partial f_{ij}}{\partial p_{j\bullet}} + (N_j - 1) \tau_{ij,j\circ} \frac{\partial f_{ij}}{\partial p_{j\circ}} & \text{for } (l, m) = (j, k) \\ \frac{\partial w_{ij}}{\partial p_{j\circ}} = \tau_{ij,j\bullet} \frac{\partial f_{ij}}{\partial p_{j\circ}} + \tau_{ij,j\circ} \left( \frac{\partial f_{ij}}{\partial p_{j\bullet}} + (N_j - 2) \frac{\partial f_{ij}}{\partial p_{j\circ}} \right) & \text{for } l = j \text{ and } m \neq k \\ \frac{\partial w_{ij}}{\partial p_l} = \tau_{ij,l} \left( \frac{\partial f_{il}}{\partial p_{l\bullet}} + (N_l - 1) \frac{\partial f_{il}}{\partial p_{l\circ}} \right) & \text{for } l \neq j. \end{cases} \quad (\text{B.3})$$

Finally, we assume that the effect of phenotype on fertility is independent of the source or destination classes of the offspring; that is, we can define  $-C = \frac{\partial f_{ij}}{\partial p_{j\bullet}} = \frac{\partial f}{\partial p_{\bullet}}$  and  $B = \frac{\partial f_{ij}}{\partial p_{j\circ}} = \frac{\partial f}{\partial p_{\circ}}$ . Combining  $C$  and  $B$  with (B.3) into equation (B.2) results in

$$\begin{aligned} \alpha \cdot \frac{d\mathbf{W}}{d\delta} \cdot \mathbf{q} &= -C \left[ \sum_{i,j=1}^K \alpha_i N_j \left( \tau_{ij,j\bullet} \overline{q_{j\bullet}^2} + (N_j - 1) \tau_{ij,j\circ} \overline{q_{j\bullet} q_{j\circ}} + \sum_{l \neq j} N_l \tau_{ij,l} \overline{q_j q_l} \right) \right] \\ &\quad B \left[ \sum_{i,j=1}^K \alpha_i N_j (N_j - 1) \left( \tau_{ij,j\bullet} \overline{q_{j\bullet}^2} + (\tau_{ij,j\bullet} + (N_j - 2) \tau_{ij,j\circ}) \overline{q_{j\bullet} q_{j\circ}} \right) + N_j \sum_{l \neq j} N_l (N_l - 1) \tau_{ij,l} \overline{q_j q_l} \right] \\ &= K_C (-C + \mathcal{R}B), \end{aligned} \quad (\text{B.4})$$

where  $K_C$  and  $K_B$  equal the first and second bracket in (B.4), respectively. Scaled relatedness is given  $\mathcal{R} = K_B/K_C$ . The terms in parentheses on the last line of (B.4) are the same as the right-hand side of equation (12), once  $C$  and  $B$  are evaluated following the assumptions in that section that the phenotype  $p$  is a parameter of some physiological/psychological mechanism that affects the underlying actions individuals choose,  $a^*$ , and the marginal effect of the action of one individual on the action of another is  $\rho$ . A detailed exposition of this analysis is given by Akçay & Van Cleve [25] where  $-C$  corresponds to  $\frac{\partial F_i}{\partial p_i}$  in equation (2) and  $B$  to  $\frac{\partial F_i}{\partial p_j}$  in equation (3) in that paper. In most biological cases of interest,  $K_C > 0$ , which results in the proportionality in equation (12). The convergence stability condition in (13) can be derived from equation (12) simply by recalling the assumption that the phenotype affects fertility but not demographic quantities contained in  $\mathcal{R}$ .

## Appendix C: Uni-directional and bi-directional reciprocity in indirect genetic effects models

In this section, we will show that: (i) assuming a uni-directional type of reciprocity where the focal individual can affect its partner’s phenotype, but not vice-versa, can lead to the expanded Hamilton’s rule presented by Queller [15, equation 1]; and (ii) bi-directional reciprocity leads to equation (17), which is equivalent to results derived by us [25] and others [11, 24]. Notation in the following equations is taken from Queller [15] and generally follows quantitative genetic approaches to social evolution [e.g., 15, 38, 53] with a particular focus on the methods in indirect genetic effect (IGE) models [24, 80, 81].

We begin with the phenotypic version of Hamilton’s rule, derived from the Price equation [47], that is equation (7) in Queller [15], which is

$$\beta_{WP.P'} + \beta_{WP'.P} \frac{\text{Cov}[G, P']}{\text{Cov}[G, P]} . \quad (\text{C.1})$$

where we use primes to denote variables associated with the social partner. Suppose that the phenotype of the social partner is defined as

$$P' = G' + \beta_{P'P}P + \epsilon' , \quad (\text{C.2})$$

which says that the phenotype of the social partner is a linear function of its average breeding value ( $G'$ ), the effect of the phenotype of the focal ( $\beta_{P'P}P$ ), and a random component with mean zero ( $\epsilon'$ , a standard assumption in quantitative genetics; see ref 82). Suppose also that the phenotype of focal individual is given by

$$P = G + \epsilon , \quad (\text{C.3})$$

which says the phenotype of the focal individual is a linear function *only* of its breeding value and a random component; thus, there is no effect of the partner’s phenotype on the focal, which embodies the assumption we believe underlies Queller [15]’s analysis. Plugging  $P$  and  $P'$  given in equations

(C.2) and (C.3) into equation (C.1) yields

$$\beta_{WP.P'} + \beta_{WP'.P} \frac{\text{Cov}[G, G' + \beta_{P'P}P + \epsilon']}{\text{Cov}[G, G + \epsilon]} = \beta_{WP.P'} + \beta_{WP'.P} (\beta_{P'P} + \beta_{G'G}) , \quad (\text{C.4})$$

which can be written as

$$-c + b(\rho + r)$$

in the notation of the main text and is equivalent to the additive combination of kin and kith components in equation (1) of Queller [15].

To include bi-directional reciprocity, we follow the analysis presented by McGlothlin *et al.* [24]; Akçay & Van Cleve [25] obtain the same results and present a comparison of the relative merits of the current approach and the IGE approach. We start by setting

$$P = G + \beta_{PP'}P' + \epsilon , \quad (\text{C.5})$$

which expresses the fact that  $P'$  also has an effect on  $P$ . Before we can plug this definition of  $P$  or the definition of  $P'$  into equation (7) of Queller [15], we must insert our definition of  $P'$  into (C.5) and our definition of  $P$  into (C.2) and solve the resulting equations; the solutions are

$$\begin{aligned} P &= \frac{G + \epsilon + \beta_{PP'}(G' + \epsilon')}{1 - \beta_{PP'}\beta_{P'P}} \\ P' &= \frac{G' + \epsilon' + \beta_{P'P}(G + \epsilon)}{1 - \beta_{PP'}\beta_{P'P}} . \end{aligned} \quad (\text{C.6})$$

These solutions can be compared to the analogous solution in McGlothlin *et al.* [24] (equation 6).

Plugging equations (C.6) into equation (7) of Queller [15] yields

$$\begin{aligned} \beta_{WP.P'} + \beta_{WP'.P} \frac{\text{Cov}\left[G, \frac{G' + \epsilon' + \beta_{P'P}(G + \epsilon)}{1 - \beta_{PP'}\beta_{P'P}}\right]}{\text{Cov}\left[G, \frac{G + \epsilon + \beta_{PP'}(G' + \epsilon')}{1 - \beta_{PP'}\beta_{P'P}}\right]} &= \beta_{WP.P'} + \beta_{WP'.P} \frac{\text{Cov}[G, G' + \epsilon' + \beta_{P'P}(G + \epsilon)]}{\text{Cov}[G, G + \epsilon + \beta_{PP'}(G' + \epsilon')]} \\ &= \beta_{WP.P'} + \beta_{WP'.P} \frac{\text{Cov}[G, G'] + \beta_{P'P} \text{Var}[G]}{\text{Var}[G] + \beta_{PP'} \text{Cov}[G, G']} \\ &= \beta_{WP.P'} + \beta_{WP'.P} \frac{\beta_{G'G} + \beta_{P'P}}{1 + \beta_{PP'}\beta_{G'G}} \end{aligned}$$

Thus, the Hamilton's rule one gets from equation (7) of Queller [15] should be

$$\beta_{WP \cdot P'} + \beta_{WP' \cdot P} \frac{\beta_{G'G} + \beta_{P'P}}{1 + \beta_{PP'}\beta_{G'G}} > 0$$

or

$$(1 + \beta_{PP'}\beta_{G'G}) \beta_{WP \cdot P'} + \beta_{WP' \cdot P} (\beta_{G'G} + \beta_{P'P}) > 0. \quad (\text{C.7})$$

Assuming that  $\beta_{P'P} = \beta_{PP'}$ , equation (C.7) becomes in our notation

$$-c(1 + \rho r) + b(r + \rho) > 0,$$

which is exactly equation (17) from our model. This is also exactly equivalent to equation (18) given by McGlothlin *et al.* [24] for pairwise interactions once the appropriate mapping between  $\rho$  and their IGE coefficient  $\psi$  is made [see Appendix A9 in ref 25].

## References

- [1] Hamilton, W. D. 1964 The genetical evolution of social behaviour. I. *J. Theor. Biol.* **7**, 1–16.
- [2] Hamilton, W. D. 1964 The genetical evolution of social behaviour. II. *J. Theor. Biol.* **7**, 17–52.
- [3] Trivers, R. L. 1971 The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57.
- [4] Axelrod, R. & Hamilton, W. D. 1981 The evolution of cooperation. *Science* **211**, 1390–1396.
- [5] Queller, D. 1985 Kinship, reciprocity and synergism in the evolution of social behaviour. *Nature* **318**, 366–367.
- [6] Nowak, M. A. & Sigmund, K. 1998 Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577.
- [7] Queller, D. C. & Strassmann, J. E. 1998 Kin selection and social insects. *Bioscience* **48**, 165–175.
- [8] Hughes, W. O. H., Oldroyd, B. P., Beekman, M. & Ratnieks, F. L. W. 2008 Ancestral monogamy shows kin selection is key to the evolution of eusociality. *Science* **320**, 1213–6. (doi:10.1126/science.1156108)
- [9] Schino, G., Di Giuseppe, F. & Visalberghi, E. 2009 Grooming, rank, and agonistic support in tufted capuchin monkeys. *Am. J. Primatol.* **71**, 101–5. (doi:10.1002/ajp.20627)
- [10] Sachs, J. L., Mueller, U. G., Wilcox, T. P. & Bull, J. J. 2004 The evolution of cooperation. *Q. Rev. Biol.* **79**, 135–160.
- [11] Lehmann, L. & Keller, L. 2006 The evolution of cooperation and altruism—a general framework and a classification of models. *J. Evol. Biol.* **19**, 1365–1376. (doi:10.1111/j.1420-9101.2006.01119.x)
- [12] Nowak, M. A. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–3. (doi:10.1126/science.1133755)
- [13] West, S. A., Griffin, A. & Gardner, A. 2007 Evolutionary explanations for cooperation. *Curr. Biol.* **17**, R661–R672.

- [14] Fletcher, J. A. & Doebeli, M. 2009 A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B* **276**, 13–9. (doi:10.1098/rspb.2008.0829)
- [15] Queller, D. C. 2011 Expanded social fitness and Hamilton’s rule for kin, kith, and kind. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 10 792–10 799. (doi:10.1073/pnas.1100298108)
- [16] Balshine-Earn, S., Neat, F. C., Reid, H. & Taborsky, M. 1998 Paying to stay or paying to breed? Field evidence for direct benefits of helping behavior in a cooperatively breeding fish. *Behav. Ecol.* **9**, 432–438.
- [17] Wilkinson, G. S. 1984 Reciprocal food sharing in the vampire bat. *Nature* **308**, 181–184.
- [18] White, C., Winans, S., White, C. & Winans, S. 2007 Cell–cell communication in the plant pathogen *Agrobacterium tumefaciens*. *Philos. Trans. R. Soc. B* **362**, 1135–1148.
- [19] Persson, O. & Öhrström, P. 1989 A new avian mating system: Ambisexual polygamy in the penduline tit *Remiz pendulinus*. *Ornis Scandinavica* **20**, 105–111.
- [20] Szentirmai, I., Székely, T. & Komdeur, J. 2007 Sexual conflict over care: antagonistic effects of clutch desertion on reproductive success of male and female penduline tits. *J. Evol. Biol.* **20**, 1739–1744.
- [21] Wesolowski, T. 1994 On the origin of parental care and the early evolution of male and female parental roles in birds. *Am. Nat.* **143**, 39–58.
- [22] Valera, F., Hoi, H. & Schleicher, B. 1997 Egg burial in penduline tits, *Remiz pendulinus*: its role in mate desertion and female polyandry. *Behav. Ecol.* **8**, 20–27.
- [23] Fletcher, J. A. & Zwick, M. 2006 Unifying the theories of inclusive fitness and reciprocal altruism. *Am. Nat.* **168**, 252–62. (doi:10.1086/506529)
- [24] McGlothlin, J. W., Moore, A. J., Wolf, J. B. & Brodie III, E. D. 2010 Interacting phenotypes and the evolutionary process. III. Social evolution. *Evolution* **64**, 2558–2574. (doi:10.1111/j.1558-5646.2010.01012.x)
- [25] Akçay, E. & Van Cleve, J. 2012 Behavioral responses in structured populations pave the way to group optimality. *Am. Nat.* **179**, 257–269. (doi:10.1086/663691)

- [26] Pepper, J. W. 2000 Relatedness in trait group models of social evolution. *J. Theor. Biol.* **206**, 355–68. (doi:10.1006/jtbi.2000.2132)
- [27] Rousset, F. & Billiard, S. 2000 A theoretical basis for measures of kin selection in subdivided populations: Finite populations and localized dispersal. *J. Evol. Biol.* **13**, 814 – 825.
- [28] Akçay, E., Van Cleve, J., Feldman, M. W. & Roughgarden, J. 2009 A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19 061–19 066. (doi:10.1073/pnas.0904357106)
- [29] André, J.-B. & Day, T. 2007 Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner’s dilemma. *J. Theor. Biol.* **247**, 11–22. (doi:10.1016/j.jtbi.2007.02.007)
- [30] Taylor, P. & Maciejewski, W. 2012 An inclusive fitness analysis of synergistic interactions in structured populations. *Proc. R. Soc. B* **279**, 4596–4603. (doi:10.1098/rspb.2012.1408)
- [31] Bulow, J. I., Geanakoplos, J. D. & Klemperer, P. D. 1985 Multimarket oligopoly: Strategic substitutes and complements. *J. Polit. Economy* **93**, pp. 488–511.
- [32] Wenseleers, T. 2006 Modelling social evolution: the relative merits and limitations of a Hamilton’s rule-based approach. *J. Evol. Biol.* **19**, 1419–1422. (doi:10.1111/j.1420-9101.2006.01144.x)
- [33] Lehmann, L. & Keller, L. 2006 Synergy, partner choice and frequency dependence: their integration into inclusive fitness theory and their interpretation in terms of direct and indirect fitness effects. *J. Evol. Biol.* **19**, 1426–1436. (doi:DOI10.1111/j.1420-9101.2006.01200.x)
- [34] Gardner, A., West, S. A. & Barton, N. H. 2007 The relation between multilocus population genetics and social evolution theory. *Am. Nat.* **169**, 207–226. (doi:10.1086/510602)
- [35] Ohtsuki, H. 2012 Does synergy rescue the evolution of cooperation? An analysis for homogeneous populations with non-overlapping generations. *J. Theor. Biol.* **307**, 20 – 28. (doi:10.1016/j.jtbi.2012.04.030)

- [36] Taylor, P. 1989 Evolutionary stability in one-parameter models under weak selection. *Theor. Popul. Biol.* **36**, 125–143.
- [37] Wild, G. & Traulsen, A. 2007 The different limits of weak selection and the evolutionary dynamics of finite populations. *J. Theor. Biol.* **247**, 382–390. (doi:10.1016/j.jtbi.2007.03.015)
- [38] Queller, D. C. 1992 Quantitative genetics, inclusive fitness, and group selection. *Am. Nat.* **139**, 540–558.
- [39] Skyrms, B. 2001 The stag hunt. *Proc. Addresses Am. Philos. Assoc.* **75**, pp. 31–41.
- [40] Taylor, C. & Nowak, M. A. 2007 Transforming the dilemma. *Evolution* **61**, 2281–2292. (doi:10.1111/j.1558-5646.2007.00196.x)
- [41] Cornforth, D. M., Sumpter, D. J. T., Brown, S. P. & Brännström, Å. 2012 Synergy and group size in microbial cooperation. *Am. Nat.* **180**, pp. 296–305. (doi:10.1086/667193)
- [42] Rousset, F. & Ronce, O. 2004 Inclusive fitness for traits affecting metapopulation demography. *Theor. Popul. Biol.* **65**, 127–141. (doi:10.1016/j.tpb.2003.09.003)
- [43] Taylor, P. D. 1990 Allele-frequency change in a class-structured population. *Am. Nat.* **135**, 95 – 106.
- [44] Charlesworth, B. 1994 *Evolution in age-structured populations*. Cambridge: Cambridge University Press, 2nd edn.
- [45] Caswell, H. 2001 *Matrix population models*. Sunderland, Mass: Sinauer Associates Sunderland.
- [46] Price, G. R. 1972 Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485–490.
- [47] Price, G. R. 1970 Selection and covariance. *Nature* **227**, 520 – 521.
- [48] Fisher, R. A. 1930 *The genetical theory of natural selection*. Oxford: The Clarendon Press.
- [49] Leslie, P. H. 1948 Some further notes on the use of matrices in population mathematics. *Biometrika* **35**, 213–245. (doi:10.1093/biomet/35.3-4.213)

- [50] Rousset, F. 2004 *Genetic structure and selection in subdivided populations*. Princeton, New Jersey: Princeton University Press.
- [51] Wright, S. 1949 The genetical structure of populations. *Ann. Eugen.* **15**, 323–354. (doi: 10.1111/j.1469-1809.1949.tb02451.x)
- [52] Wright, S. 1931 Evolution in mendelian populations. *Genetics* **16**, 97–159.
- [53] Frank, S. A. 1997 The Price Equation, Fisher’s fundamental theorem, kin selection, and causal analysis. *Evolution* **51**, 1712 – 1729.
- [54] Hamilton, W. 1972 Altruism and related phenomena, mainly in social insects. *Annu. Rev. Ecol. Syst.* **3**, 193–232.
- [55] Queller, D. C. 1992 A general model for kin selection. *Evolution* **46**, 376 – 380.
- [56] Queller, D. C. 1994 Genetic relatedness in viscous populations. *Evol. Ecol.* **8**, 70–73. (doi: 10.1007/BF01237667)
- [57] Lehmann, L. & Rousset, F. 2010 How life history and demography promote or inhibit the evolution of helping behaviours. *Philos. Trans. R. Soc. B* **365**, 2599–2617. (doi:10.1098/rstb.2010.0138)
- [58] smith, j., Van Dyken, J. D. & Zee, P. C. 2010 A generalization of Hamilton’s rule for the evolution of microbial cooperation. *Science* **328**, 1700–3. (doi:10.1126/science.1189675)
- [59] Eshel, I. & Motro, U. 1981 Kin selection and strong evolutionary stability of mutual help. *Theor. Popul. Biol.* **19**, 420–433.
- [60] Eshel, I. 1983 Evolutionary and continuous stability. *J. Theor. Biol.* **103**, 99 – 112.
- [61] Wakano, J. Y. & Lehmann, L. 2012 Evolutionary and convergence stability for continuous phenotypes in finite populations derived from two-allele models. *J. Theor. Biol.* **310**, 206 – 215. (doi:10.1016/j.jtbi.2012.06.036)
- [62] Bijma, P. & Wade, M. J. 2008 The joint effects of kin, multilevel selection and indirect genetic effects on response to genetic selection. *J. Evol. Biol.* **21**, 1175–1188. (doi:10.1111/j.1420-9101.2008.01550.x)

- [63] Maynard Smith, J. & Szathmáry, E. 1995 *The major transitions in evolution*. Oxford: W.H. Freeman.
- [64] Lande, R. & Arnold, S. J. 1983 The measurement of selection on correlated characters. *Evolution* **37**, pp. 1210–1226.
- [65] Heisler, I. L. & Damuth, J. 1987 A method for analyzing selection in hierarchically structured populations. *Am. Nat.* **130**, 582–602.
- [66] Goodnight, C. J., Schwartz, J. M. & Stevens, L. 1992 Contextual analysis of models of group selection, soft selection, hard selection, and the evolution of altruism. *Am. Nat.* **140**, pp. 743–761.
- [67] Price, G. R. 1972 Fisher’s ‘fundamental theorem’ made clear. *Ann. Hum. Genet.* **36**, 129–40.
- [68] Ewens, W. J. 1989 An interpretation and proof of the fundamental theorem of natural selection. *Theor. Popul. Biol.* **36**, 167–180.
- [69] Heywood, J. S. 2005 An exact form of the breeder’s equation for the evolution of a quantitative trait under natural selection. *Evolution* **59**, 2287–2298.
- [70] Moore, A. J., Haynes, K. F., Preziosi, R. F. & Moore, P. J. 2002 The evolution of interacting phenotypes: genetics and evolution of social dominance. *Am. Nat.* **160 Suppl 6**, S186–97. (doi:10.1086/342899)
- [71] Petfield, D., Chenoweth, S. F., Rundle, H. D. & Blows, M. W. 2005 Genetic variance in female condition predicts indirect genetic variance in male sexual display traits. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 6045–50. (doi:10.1073/pnas.0409378102)
- [72] Mutic, J. J. & Wolf, J. B. 2007 Indirect genetic effects from ecological interactions in *Arabidopsis thaliana*. *Mol. Ecol.* **16**, 2371–2381.
- [73] Ellen, E. D., Visscher, J., Arendonk, J. A. M. v. & Bijma, P. 2008 Survival of laying hens: Genetic parameters for direct and associative effects in three purebred layer lines. *Poult. Sci.* **87**, 233–239. (doi:10.3382/ps.2007-00374)

- [74] Danielson-François, A. M., Zhou, Y. & Greenfield, M. D. 2009 Indirect genetic effects and the lek paradox: inter-genotypic competition may strengthen genotype x environment interactions and conserve genetic variance. *Genetica* **136**, 27–36. (doi:10.1007/s10709-008-9297-z)
- [75] Bleakley, B. H. & Brodie, III, E. D. 2009 Indirect genetic effects influence antipredator behavior in guppies: estimates of the coefficient of interaction  $\psi$  and the inheritance of reciprocity. *Evolution* **63**, 1796–806. (doi:10.1111/j.1558-5646.2009.00672.x)
- [76] Teplitsky, C., Mills, J. A., Yarrall, J. W. & Merilä, J. 2010 Indirect genetic effects in a sex-limited trait: the case of breeding time in red-billed gulls. *J. Evol. Biol.* **23**, 935–944. (doi:10.1111/j.1420-9101.2010.01959.x)
- [77] Frère, C. H., Krützen, M., Mann, J., Connor, R. C., Bejder, L. & Sherwin, W. B. 2010 Social and genetic interactions drive fitness variation in a free-living dolphin population. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19 949–19 954. (doi:10.1073/pnas.1007997107)
- [78] Wilson, A. J., Morrissey, M. B., Adams, M. J., Walling, C. A., Guinness, F. E., Pemberton, J. M., Clutton-Brock, T. H. & Kruuk, L. E. B. 2011 Indirect genetics effects and evolutionary constraint: an analysis of social dominance in red deer, *Cervus elaphus*. *J. Evol. Biol.* **24**, 772–83. (doi:10.1111/j.1420-9101.2010.02212.x)
- [79] Hamilton, I. M. & Ligocki, I. Y. 2012 The extended personality: indirect effects of behavioural syndromes on the behaviour of others in a group-living cichlid. *Anim. Behav.* **84**, 659–664. (doi:10.1016/j.anbehav.2012.06.022)
- [80] Moore, A. J., Brodie, III, E. D. & Wolf, J. B. 1997 Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution* **51**, 1352–1362.
- [81] Wolf, J. B., Brodie, III, E. D. & Moore, A. J. 1999 Interacting phenotypes and the evolutionary process. II. Selection resulting from social interactions. *Am. Nat.* **153**, 254–266. (doi:10.1086/303168)
- [82] Lynch, M. & Walsh, B. 1998 *Genetics and analysis of quantitative traits*. Sunderland, Mass.: Sinauer.

- [83] Queller, D. C., Ponte, E., Bozzaro, S. & Strassmann, J. E. 2003 Single-gene greenbeard effects in the social amoeba *Dictyostelium discoideum*. *Science* **299**, 105–106.
- [84] Clutton-Brock, T. H. 1991 *The evolution of parental care*. Princeton, NJ: Princeton University Press.
- [85] Holekamp, K. E., Smith, J. E., Strelhoff, C. C., Van Horn, R. C. & Watts, H. E. 2012 Society, demography and genetic structure in the spotted hyena. *Mol. Ecol.* **21**, 613–632. (doi:10.1111/j.1365-294X.2011.05240.x)
- [86] Apicella, C. L., Marlowe, F. W., Fowler, J. H. & Christakis, N. A. 2012 Social networks and cooperation in hunter-gatherers. *Nature* **481**, 497–501.

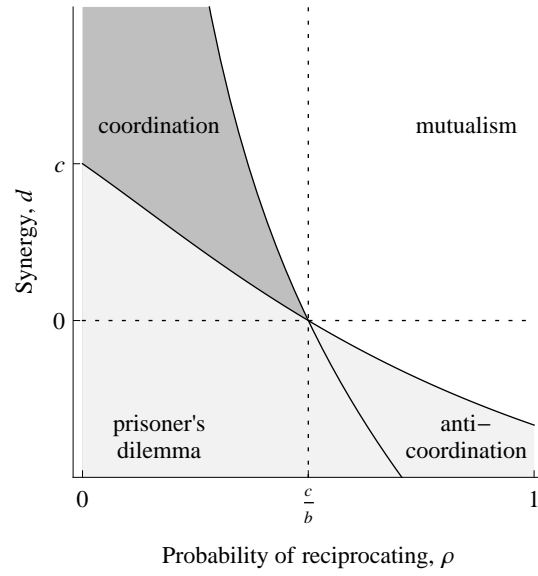


Figure 1: The different possibilities for how a prisoner's dilemma game with reciprocity and non-additive payoffs can get transformed, depending on the probabilities of reciprocation  $\rho$  and the non-additive payoff or synergy component  $d$ .